

Expansion of the Genomic Encyclopedia of Bacteria and Archaea

Christian Rinke¹, Alex Sczyrba¹, Stephanie Malfatti¹, Janey Lee¹, Jan-Fang Cheng¹, Ramunas Stepanauskas², Jonathan A. Eisen^{1,3}, Steven Hallam⁴; William P. Inskeep⁵; Brian P. Hedlund⁶; Stefan M. Sievert⁷; Wen-Tso Liu⁸; George Tsiamis⁹; Philip Hugenholtz¹⁰; Tanja Woyke¹

¹ DOE Joint Genome Institute, Walnut Creek, California, United States of America; ² Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine, United States of America; ³ Department of Evolution and Ecology, University of California Davis, Davis, United States of America, ⁴ Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada; ⁵ Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, United States of America; ⁶ School of Life Sciences, University of Nevada, Las Vegas, Nevada, United States of America; ⁷ Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, United States of America; ⁸ Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America; ⁹ Department of Environmental and Natural Resources Management, University of Ioannina, Agrinio, Greece; ¹⁰ Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Australia.

March 2011

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Christian Rinke¹, Alex Sczyrba¹, Stephanie Malfatti¹, Janey Lee¹, Jan-Fang Cheng¹,
Ramunas Stepanauskas², Jonathan A. Eisen^{1,3}, Steven Hallam⁴; William P. Inskeep⁵;
Brian P. Hedlund⁶; Stefan M. Sievert⁷; Wen-Tso Liu⁸; George Tsiamis⁹; Philip
Hugenholtz¹⁰; Tanja Woyke¹

¹ DOE Joint Genome Institute, Walnut Creek, California, United States of America; ² Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine, United States of America; ³ Department of Evolution and Ecology, University of California Davis, Davis, United States of America; ⁴ Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada; ⁵ Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, United States of America; ⁶ School of Life Sciences, University of Nevada, Las Vegas, Nevada, United States of America; ⁷ Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, United States of America; ⁸ Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America; ⁹ Department of Environmental and Natural Resources Management, University of Ioannina, Agrinio, Greece; ¹⁰ Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Australia.

To date the vast majority of bacterial and archaeal genomes sequenced are of rather limited phylogenetic diversity as they were chosen based on their physiology and/ or medical importance. The Genomic Encyclopedia of Bacteria and Archaea (GEBA) project (Wu et al. 2009) is aimed to systematically filling the gaps of the tree of life with phylogenetically diverse reference genomes. However more than 99% of microorganisms elude current culturing attempts, severely limiting the ability to recover complete or even partial genomes of these largely mysterious species. These limitations gave rise to the GEBA uncultured project. Here we propose to use single cell genomics to massively expand the Genomic Encyclopedia of Bacteria and Archaea by targeting 80 single cell representatives of uncultured candidate phyla which have no or very few cultured representatives. Generating these reference genomes of uncultured microbes will dramatically increase the discovery rate of novel protein families and biological functions, shed light on the numerous underrepresented phyla that likely play important roles in the environment, and will assist in improving the reconstruction of the evolutionary history of *Bacteria* and *Archaea*. Moreover, these data will improve our ability to interpret metagenomics sequence data from diverse environments, which will be of tremendous value for microbial ecology and evolutionary studies to come.

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No.

DE-AC02-05CH11231